DOCUMENT RESUME

ED 205 179 IR 009 482

TITLE	An Introduction to 3RIP--An Interactive Text Data Base System.
INSTITUTION	Royal Inst. of Tech., Stockholm (Sweden). Library.
REPORT NO	TRITA-LIB-4068
PUB DATE	Mar 81
NOTE	31p.: For related document, see ED 196 437.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Databases; Editing; *Information Retrieval; Information Systems; *Online Systems; *Search Strategies
IDENTIFIERS	*3RIP Search System; Command Language; Text Editors

ABSTRACT

This descriptive guide provides introductory information geared to aid both the user and the information manager in the use of 3RIP, an online text database system consisting of three components: (1) a retrieval system used to locate and display records, (2) a file handler used to read records from files and to write records onto files, and (3) a text editor used to create and modify records. Chapters describe the system in general, 3RIP's menu-based dialogue for novices and occasional users, and 3RIP's standard command vocabulary. A search sample is displayed, and the managerial aspects of 3RIP are examined. An appendix contains some descriptions of 3RIP's demonstration search files. (FM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

REPORT TRITA-LIB-4068

MARCH 1981

ED205179

AN INTRODUCTION TO 3RIP -

AN INTERACTIVE TEXT DATA BASE SYSTEM

Paralog AB
Stockholm, Sweden

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stephan Schwarz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FR009482

EDITOR'S PREFACE

The interactive text data base system 3RIP is the result of a development project that started in 1973 at the Royal Institute of Technology Library. Since 1976 the development of 3RIP has been done by the consulting company Paralog AB which was formed by members of the 3RIP development group.

The present introductory description of 3RIP was originally issued by Paralog AB in July 1980, and it is now re-issued in the Stockholm Papers in Library and Information Science series as a sample description of a contemporary system for online information retrieval. As such it is of a general interest to the information science community.

The work of the 3RIP development group, and later Paralog AB, has been published in a number of reports in this series. Of these the following are still relevant:

Löfström, M. et al. "3RIP - Data Structures . . .
TRITA-LIB-4061, June 1977

Löfström, M. "3RIP: Evaluation . . .
TRITA-LIB-4064, March 1979

Larsson, R. "On Implementing . . .
TRITA-LIB-4067, April 1980.

Since 3RIP, like most comparable systems, may be changed in response to user needs, the present text should not be considered a manual for 3RIP but a description of the system as it stood in 1980. Details about the current status of 3RIP can be obtained from Paralog AB or from any of the 3RIP installation sites.

Table of contents
==========

Foreword 7

3RIP in General 8

Dialogue and Help 10

3RIP's Command Vocabulary

- Introduction 11
- Search commands 12
- Search term patterns 12
- Auxiliary search system commands 13
- File handling commands 14
- Editing commands 14
- Editing arguments 15
- Modifiers 15
- Types and data base names 16
- Auxiliary symbols 16

A Major Search Example 17

Manageriai Aspects

- System requirements 24
- Search file systems 24
- The files in a search file system 25
- Adding new records to a data base 26
- Some 3RIP data bases 27

Appendix

- A description of WONDER 29
- Output formats for WONDER 31
- A description of MIRROR and POEMS 33

Foreword

To the user:

This introductory text focuses on how 3RIP is used for searching in text data bases. All examples are taken from search sessions with 3RIP's demonstration data bases WONDER, MIRROR, and POEMS. Fully to appreciate the examples some knowledge of the data bases is necessary. The perusal of the descriptions of the data bases in the Appendix is therefore strongly recommended.

To the 3RIP manager:

Focusing on searching as it does, this introductory text makes no claim to functioning as a complete guide to data base management with 3RIP. The lists of editing and file handling commands does, however, illustrate the ease and flexibility of the tools 3RIP offers to that purpose. It is also to be hoped that the section on Managerial Aspects will answer some of your questions on the economy and efficiency of 3RIP.

We welcome further enquiries.

Paralog
Box 2284
S-10317 Stockholm
Sweden

Telephone: Int'l + 46 (8) 144190

3RIP in General

The INTERACTIVE TEXT DATA BASE SYSTEM 3RIP is owned, developed, and maintained by Paralog. 3RIP consists of:

a RETRIEVAL SYSTEM used to locate RECORDS (texts, documents) in one, or in parallel in several, DATA BASES and to display the records in different ways;
a FILE HANDLER used to read records from FILES and data bases and to write records onto files;
a TEXT EDITOR used to create and modify records.

Data bases as well as files consist of records, but the retrieval system can only reach the records of the data bases.

The formal restriction on the size of a data base is that it should not exceed four million records, but for reasons of expediency and security no more than half a million records should be put in one data base. Parallel and simultaneous searches are possible in up to 30 data bases, in what is called a search file system. Any number of such search file systems may be active at an installation. The records may be of any length as long as they can be read into core in their entirety, but 3RIP works best if the majority of the records are not more than 2500 characters long, or roughly one page of text.

Up to 30 different search types may be employed for a data base. This means that the contents of a record may be classified into 30 types of information, e.g. journal, date of publication, author, title, keyword descriptors, an abstract, etc. for a bibliographical data base. A search order such as:

```
* AU=John Doe AND ABS=Western Culture
```

would then typically identify records where John Doe is given as author and the phrase Western Culture appears somewhere in the abstract.

Every word in a record may be made retrievable, or only phrases, or every word for some search types and only phrases for others. Pattern matching is also possible:

```
* colo:r
```

will identify COLOUR as well as COLOR and

```
* #west#
```

finds WEST, SOUTHWEST, WESTERN, etc.

The retrieved records may be displayed on your terminal, sent to the lineprinter, or written in a disk file for storage or further editing. Any number of output formats are allowed for the data bases. The format decides what parts of the records found should be displayed and allows for the setting of margins, indenting of lines, adding of blank lines, and insertion of explanatory texts.

Most users of 3RIP will probably be interested in the retrieval system only and need not bother about the file handler and editor. As such a user you need a password to enter 3RIP, giving access to the data bases you want to search, information on how the data bases are organized and what they contain, and some knowledge of 3RIP's facilities for searching and presenting search results.

Extensive help texts are available on line as you run 3RIP, in fact the whole of this document is but a part of the help texts. There are also 3RIP manuals

for searching with 3RIP and for managing 3RIP data bases.

The help texts (in particular the examples) make heavy use of 3RIP's standard demonstration search file system with three data bases: WONDER (Alice in Wonderland), MIRROR (Through the Looking-Glass), and POEMS (the poems from the two books). You should be able to try this system out by entering 3RIP with the password LEWIS; if you are a beginner you are advised to do so.

The help texts on the file handler, the editor, the individual file handling and editing commands, and the establishing of a data base are intended mainly for the use of 3RIP-managers responsible for the running of a 3RIP system and for others who wish to establish private data bases or other 3RIP-files.

Dialogue and Help

In addition to the standard 3RIP dialogue and command vocabulary, discussed in this document, there is a menu-based 3RIP dialogue available, which might be preferred by the novice or the occasional user. In this form 3RIP presents a menu of commands to be chosen from each time an order has been executed. The menu-dialogue may be tried on 3RIP's demonstration data bases by entering 3RIP with the password CARROLL.

Standard 3RIP prompts you to enter a new order by writing an * in the leftmost position of a new line on your terminal. The order must be entered after the asterisk. The order will be processed when you depress the <RETURN> key.

Once your order has been received 3RIP checks its syntax. Errors in an order cause its processing to be aborted with an error message, otherwise it is executed and an asterisk is printed on a new line showing 3RIP to be ready for the next order.

If you are unsure of the syntax of the command you wish to use you will always get help by entering a help order, for example:

```
* HELP PRINT
```

Instead of the HELP-command you may use the help string ?? to reach the same result, in this case:

```
* PRINT??
```

The help text for a command normally illustrates the use of the command and always gives its formal syntax.

For most commands it is not necessary to use the full form (complete with various arguments) directly. You may give just the command itself in an order. 3RIP will then give directions on how to continue. For example:

```
* PRINT
```

The PRINT-command needs to know what search results to display, thus if this information is missing in the original order 3RIP will ask for it, at the same time indicating a default answer: the last set created by a search order. Thus:

```
Set: /LAST/
```

When you have answered this question 3RIP will go on to ask how many records to display from each data base, as this, too, may be given in the full form of the command:

```
Number of WONDER records: /1/

(etc. etc.)
```

3RIP's Command Vocabulary

Introduction

3RIP recognizes just under a hundred "words" and single characters as belonging to its vocabulary.

First there are the commands, modifiers, arguments, search types, and data base names you use in your orders to 3RIP. You can get information about any one of those by giving an order just containing its name and the help string ??. They can be abbreviated in orders. Just type a few of the initial letters of the name of a command and then the abbreviation character .; e.g. PRO. for the command PROTECT in 3RIP' standard vocabulary. If you happen to abbreviate the command into ambiguity 3RIP will list the possibilities and ask which one you mean; for example, PR. in an order causes 3RIP to ask you to choose between PRINT and PROTECT, which both begin in PR. You can also change the names of these symbols with the command CHANGE.

Secondly there are a number of symbols fulfilling important syntactical functions. Some of them, such as = ( ) and the editor control characters <ESC> and @, can be changed with the command CHANGE. Some cannot be changed during a 3RIP-session (your manager could, however, arrange for another password, which would define a different set of "names" for these elements). The most important of these are: the three pattern forming characters # : !; the help string ??; the abbreviation character .; the general string delimiter "; and the display fill character_.

The string delimiter " enables you to make 3RIP interpret an element from the command vocabulary as a search string rather than as a command. For example: in 3RIP's standard vocabulary AND is a command but surrounded by string delimiters, "AND", it will be accepted as a search string.

Under 3RIP's standard character set no distinction is made between upper and lower case characters in orders, and characters other than letters, digits, and the three pattern forming characters # : ! are treated as indistinguishable. Initial and final sequences of indistinguishable characters are deleted, and one display fill character is substituted for any internal sequence of such characters. The display fill character is also used to tie together the words in retrieval strings whose search types are of the keyword variety.

For example: the name "Edwin, Earl of Mercia" belongs to the keyword search type PERS in 3RIP's demonstration data bases. It can be retrieved with a search order:

```
* pers=Edwin, Earl of Mercia
```

but a later LIST-order will show that 3RIP normalized this to:

```
PERS=EDWIN_EARL_OF_MERCIA
```

Search commands

There is no special "find"-command: normally 3RIP interprets an order with no command as a search order, locating records containing the submitted string. In 3RIP the following commands are used to combine pairs of search expressions.

AND	finds records where the expressions to its left and its right BOTH occur
NOT	finds records where the expression to its left but NOT the one to its right occur
OR	finds records where at least one of the expressions to its right and left occcur
W1	finds records where the expression to its right follows immediately on the one on its left
W2	finds records where the expression on its right follows within two words of the one on its left
W3	analogous to W2
W4	analogous to W2
W5	analogous to W2
WS	finds records where the expressions it combines occur in the same sentence
WP	finds records where the expressions it combines occur in the same paragraph

Search term patterns

A simple search expression can only be used to locate words or keywords that match it exactly. However, if PATTERN EXPRESSIONs are used partial matches can be located as well.

Pattern expressions and simple search expressions look alike except that the character strings of the former contain at least one of the three PATTERN FORMING CHARACTERs # : and !.

\# represents an arbitrary character string (including the empty string). Thus:

- LOCK# represents any string that begins with LOCK such as lock, locks, or locked;
- #LOCK represents any string that ends with LOCK such as lock, flock, clock, or o'clock;
- #LOCK# represents any string that contains the string LOCK such as lock, locks, locked, unlocking, flock, clock, and o'clock.

: represents ONE arbitrary character or NO character. Thus:

- LOCK: represents any string of four or five characters beginning with LOCK such as lock or locks, but not locked;
- ::LOCK represents any string of four to six characters ending with LOCK such as lock, flock, or clock, but not o'clock;

! represents ONE arbitrary character. Thus:

- LOCK! represents any string of five characters beginning with LOCK such as locks but neither lock nor locked;
- !!!LOCK represents any string of seven characters ending in LOCK such as o'clock but not lock or flock or clock;
- ADVI!E represents both the forms advice and advise.

The 3RIP-manager may choose to set up a data base so that pattern expressions are illegal, or so that, although pattern expressions are allowed, they must not begin with pattern forming characters. For details about an individual data base, give the STATUS-command for the data base.

When a pattern expression is encountered in a search order 3RIP matches it against a VOCABULARY FILE containing all the TERMs (i.e. the DIFFERENT words, keywords, or keyword phrases) in the available data bases. What happens then depends on which of two selection modes 3RIP is currently running in:

- ALL: all terms in the vocabulary file matching the pattern expression are selected and combined in a set; there is a limit to the number of terms allowed to combine automatically, 20 in the standard 3RIP-version, if there are more matches then 3RIP will select none and prompt you to be more specific or to try in the other mode;
- SELECT: the matching terms of the file are presented ten at a time for your inspection and/or selection.

The selection mode of the system may be overridden for a particular pattern expression in an order by preceding the expression with one of the symbols SELECT or ALL.

When you enter the standard version 3RIP is in ALL mode. A general change from one mode to the other may be made with one of the orders:
*CHANGE ALL TO SELECT.
*CHANGE SELECT TO ALL

In response to the directions accompanying the term lists when 3RIP is in SELECT mode the following 3RIP-symbols may be used:

ALL	select all the displayed terms and associate them with the pattern expression
GROUP	form a set of the terms indicated after this symbol without associating the set with the pattern expression
SEPARATE	form a separate set for each of the terms indicated after this symbol without associating the sets with the pattern expression
SPECIFY	specify the distribution over types and data bases for the terms indicated after this symbol
MORE	display the next ten terms from the vocabulary file
BACK	display the preceding ten terms from the vocabulary file
LEAVE	form a set for all hitherto selected terms associated with the pattern expression and stop selection and vocabulary inspection

Auxiliary search system commands

An order with one of the commands below must not contain any other command:

PRINT	Present all or some of the records of a set
PURGE	Delete one or more sets
LIST	List a set directory giving the sizes of sets and the orders used to create them
NAME	Assign a name to a set
FORM	Change output format for data base records
OPEN	Open one or more data bases for searching and/or vocabulary inspection
CLOSE	Close one or more data bases (partially) for searching and/or

	close them for vocabulary inspection
TRACE	Display a set directory for the order sequence that was used to create a set
SAVE	Create an order record containing the order sequence that was used to create a set
ROUTINE	Create an order record (with or without search orders)
RUN	Execute the order sequence saved in an order record
STATUS	Give status information about the search system
STATISTICS	Compute simple statistics for a value set (one with numeric values)
RESTRICT	Create a new value set by applying value restrictions to an old set
CHANGE	Change default type, selection mode, or the names of individual 3RIP-symbols
GETSYM	Get a new command vocabulary for 3RIP
PUTSYM	Save the current command vocabulary in a symbol record
EXIT	Leave 3RIP
RESTART	Resume an interrupted search
HELP	Command equivalent to ??, which is to be preferred

File handling commands

GET	Read a record from a file or a data base
PUT	Write a record onto a file
DIR	Give name, size, version, protection, etc. for each of the indicated records of a file
TYPE	Output the indicated records of a file or a data base
X	Delete a record from a file
REN	Rename a record in a file
PROTECT	Protect a record against unauthorized access
FORM	Change the format for the output from TYPE
INFILE	Define a default input file
OUTFILE	Define a default output file
PACK	Reorganize a file deleting all old record versions
CHANGE	Change the names of individual 3RIP-symbols
GETSYM	Get a new command vocabulary for 3RIP
PUTSYM	Save the current command vocabulary in a symbol record
ROUTINE	Create an order record
RUN	Execute the order sequence saved in an order record
EXIT	Leave 3RIP
HELP	Command equivalent to ??, which is to be preferred

Editing commands

LOC	Locate a string
S	Substitute a string for a string; a subfield; a text interval or variable; or a given number of characters, words, sentences, or lines
I	Insert a string
X	Delete a record; a field; a subfield; a text interval or variable; or a given number of characters, words, sentences, or lines
D	Show the contents of a record; a field; a subfield; a text interval or variable; or a given number of characters, words, sentences, or lines
SC	Assign a name to the current cursor position

TC	Move the cursor to a named position
CC	Indicate the cursor position
FORM	Change the format used by D
MAKE	Initialize a new record
INFORM	Fetch a data entry format
CREATE	Create a record using controlled data entry
STATUS	Give status information about the editing system
GET	Read a record from a file or a data base
PUT	Write a record onto a file
REN	Rename a record
CHANGE	Change editor control characters or the names of individual 3RIP-symbols
GETSYM	Get a new command vocabulary for 3RIP
PUTSYM	Save the current command vocabulary in a symbol record
ROUTINE	Create an order record
RUN	Execute the order sequence saved in an order record
EXIT	Leave 3RIP
HELP	Command equivalent to ??, which is to be preferred

The name of a text interval shall be @Y or @Z and that of a text variable shall be @A, @B, ..., or @X. Some of the commands require the use of a delimiter: <ESC>. This and the character @ are the editor control characters referred to in connexion with the command CHANGE above.

Editing arguments	Referring to
C <n>	n characters forward (or backward if n negative)
W <n>	n words forward (or backward if n negative)
SEN <n>	n sentences forward (or backward if n negative)
L <n>	n lines forward (or backward if n negative)
SF <n>	subfield n
F <n> <m>	subfield m of field n
R <record name>	the given record
<text variable>	the given variable (named @A, @B,..., @X)
<text interval>	the given interval (named @Y or @Z)

Modifiers

REVERSED	in PRINT-orders (selects records for output in descending record number order from the current set)
SORTED	in PRINT- and TYPE-orders (sorts the selected records before output)
LPT	in PRINT-, TYPE-, STATUS-, and DIR-orders (sends the output to a line-printer instead of the terminal)
TEXTFILE	in PRINT-, TYPE-, STATUS-, and DIR-orders (appends the output to an ordinary ASCII text file)
TFORM	in PRINT- and TYPE-orders (appends the output to an ordinary ASCII text file in 3RIP's input format, TFORM)
FILE	in PRINT- and TYPE-orders (appends the output to a 3RIP-file)
BACKUP	in TYPE- and GET-orders (picks the preceding generation of the selected records, not the current one)
ORIGINAL	in TYPE- and GET-orders (picks the oldest generation of the selected records, not the current one)

FORCING	in TYPE-orders (asks the user for the correct password before displaying protected 3RIP-records)
STRUCTURE	in FORM-, X-, and I-orders (chooses an output format showing the physical structure of 3RIP-records with the first type of order; changes the structure of a record with the last two types of orders)
DUMP	in FORM-orders (chooses an output format showing the physical structure of 3RIP-records while giving the octal ASCII-code for all non-printable characters)
VOCABULARY	in CLOSE- and OPEN-orders (limits the closing or opening to the vocabulary file)
SUMMARY	in DIR-orders (restricts the output of directory data to summary line)
SEARCH	in STATUS-orders (gives information about the search system)
EDIT	in STATUS-orders (gives information about the editing system)

Types and data base names

In the demonstration search system available under the password LEWIS the following data bases and search types are available:

```
Default type:       TEXT
Free text types:    TEXT
Keyword types:      PERS CH PT

Data bases:         WONDER MIRROR POEMS
```

The STATUS-command gives more information for the individual data bases.

Auxiliary symbols

The 3RIP-symbol = binds a type to a character string thus forming a simple search expression. The symbol may only be given in search expressions and in set names, and in both cases it shall have the 3RIP-symbol for a type (or the corresponding number between 2 and 31) on its left and a character string on its right.

The 3RIP-symbols (and) may be used to alter the evaluation order for search expressions containing several search commands.

The 3RIP-symbols TO and FROM may be used to set a date limit in CLOSE- and DIR-orders. They can also be used in responses to some directions and to set record number limits for individual data bases in CLOSE-order.

The 3RIP-symbol TO may also be used in PURGE- and LIST-orders to indicate sequences of sets, in RESTRICT-orders to define value restrictions, and in CHANGE-orders to alter defaults.

A Major Search Example

The following example illustrates a session with the Carroll data bases under the password LEWIS.

```
.r 3RIP

Copyright (c) 1979, PARALOG AB
Date: 80-02-19  Time: 11:40:58
Password:
*
```

COMMENTS:
The monitor command "r 3RIP" is the usual way to start 3RIP. Your computer service may require another command. The entered password is not echoed; LEWIS was used here and should work on your computer too. It gives access to 3RIP's standard demonstration data bases. Other passwords will make other data bases available. The prompter * signals that 3RIP is waiting for an order.

```
* STATUS SEARCH
              Open for      Vocabulary
Data base     Retrieval     open
---------     ---------     ----------
WONDER        yes           yes
MIRROR        yes           yes
POEMS         yes           yes

Search types
------------
Default:      TEXT
Free text:    TEXT
Keyword:      PERS CH PT

Pattern forming characters
--------------------------
#  any string or the empty string
:  any one character or none
!  any one character

Current output format: 2
Number of sets used:   0
```

COMMENTS:
We ask for the status of the search system and are told the names of the available data bases, whether they are open for retrieval and whether their vocabulary is open for the use of search term patterns, what search types may be used, what pattern forming characters to use, the number of the data base output format, and the number of sets so far created.

```
* STATUS wonder
===== Data base: WONDER =====
Choose one of these alternatives:
        0.  Leave status mode
        1.  Describe the data base WONDER
        2.  Describe the data base output formats
        3.  Describe the search types
        4.  Show the date/record number relations
        5.  Show update information
Alternative number? /0/ 1
        The data base WONDER contains ...
```

COMMENTS:
It is essential that you have a fair knowledge of the basic characteristics of the data bases you wish to search. The command STATUS followed by individual data base names or by the 3RIP-symbol ALL allows you to gain that knowledge (provided your 3RIP-manager has written the necessary texts). The order * STATUS LPT ALL will print all the available information about the search system and the individual data bases on your line printer (if you use TEXTFILE instead of LPT then it will be written in a disk file). You will find part of the status information for the data bases WONDER, MIRROR, and POEMS in the Appendix of this text.

```
* present OR presents
Set 1: 12 hits in 6 records
* birthday W1 1
Set 2: 3 hits in 3 records
* un-birthday W1 1
Set 3: 3 hits in 3 records
* 3 NOT 2
Set 4: 1 hit in 1 record
* (un-birthday NOT birthday) W1 (present OR presents)
Set 5: 1 hit in 1 record
* 4 NOT 5
Set 6: 0 hits
```

COMMENTS:
Set 1 refers to the six records where at least one of the two words "present" or "presents" occurs as search type TEXT. Of these records three are such that the word "birthday" immediately precedes one of the two words. There are also (set 3) three records where one of the two words from set 1 is immediately preceded by the word "un-birthday". Sets 2 and 3 are not identical, however, as set 4 shows. Another way to create a set referring to the same records as set 4 is to cram it all into one order (creating set 5). That sets 4 and 5 are identical is shown by set 6. The 3RIP-symbols (and) in the order for set 5 are necessary because of the priority rules for the search commands.

```
* SELECT present#
Type: TEXT
 1 <7>          PRESENT
 2 <1>          PRESENTED
 3 <2>          PRESENTLY
 4 <5>          PRESENTS
No more hits
Action? /LEAVE/ 1 4
Action? /LEAVE/
Set 7: 12 hits
* present#
Set 7: 12 hits
* PURGE 7
* present#
Set 7: 15 hits
```

COMMENTS:
We are in selection mode ALL by default, but use SELECT to get a list of terms matching the pattern present# (the number of occurences are given between <>-symbols). Responding first with the term numbers 1 and 4 and then with the default we cause 3RIP to form set 7, identical to set 1. The next order * present# refers us back to set 7, for a pattern expression automatically names the set it creates. Purging set 7 (which also deletes the name) and trying again we get a new set 7 (and a naming set 8). It refers not only to "present" and "presents" but also to "presented" and "presently".

```
* grin WS cat AND pers=cheshire cat
Set 9: 2 hits in 2 records
* 9 AND pers=sp_cheshire cat
Set 10: 1 hit in 1 record
* TRACE 2
Set     Size    Order
  1       12    TEXT=PRESENT OR PRESENTS
  2        3    TEXT=BIRTHDAY W1 1
* SAVE 2 birthpres order
```

COMMENTS:
Set 9 refers to records where the words "grin" and "cat" (by default taken as search type TEXT) occur in the same sentence and where the name "Cheshire Cat" appears as the search type PERS. The only difference in set 10 is that the Cheshire Cat must appear as PERS with the prefix sp_ (for speaker). The order * TRACE 2 lists all sets involved in the creation of set 2. The last order causes 3RIP to save the orders involved in the creation of set 2 in a record called BIRTHPRES in a file called ORDER. If the given file does not exist then it will be created. If the given record already exists then it will be replaced.

```
* CHANGE
3RIP-symbol: OR
Substitute: +
3RIP-symbol: NOT
Substitute: -
3RIP-symbol:
* RUN birthpres order
TEXT=PRESENT + PRESENTS
Set 11: 12 hits in 6 records
TEXT=BIRTHDAY W1 11
Set 12: 3 hits in 3 records
```

COMMENTS:
The order * CHANGE causes 3RIP to ask for the name of a symbol to change and, when that has been given, for the new name. Only when the first question is answered by <RETURN> does 3RIP give back the prompter *. Here we have changed the symbols OR and NOT to + and -. The RUN-command (executing the order record saved earlier) consequently types + instead of OR. Sets 11 and 12 are identical to sets 1 and 2.

```
* PUTSYM boole symbol
* wabe
Set 13: 4 hits
* CLOSE wonder mirror
* wabe
Set 14: 2 hits
* 'twas brillig "and" the slithy toves
Set 15: 2 hits in 2 records
```

COMMENTS:
The PUTSYM-command saves the present symbol set (i.e. including + and - for OR and NOT) in the record BOOLE in the file SYMBOL. Set 13 refers to occurrences of "wabe" in all three data bases. We close MIRROR and WONDER. A repeated search for "wabe" finds the two occurrences in the remaining data base POEMS. In the last order the word AND must be enclosed by "-characters if it is a 3RIP-symbol and we want it to be interpreted as a search string and not as a command.

```
* PRINT 15
Record: 37    Data base: POEMS
        This record is from:
           Chapter 1 : Looking-Glass House
           Poem:           Jabberwocky
        The poem occurs ...
(here follows more "overhead" information)
Skip to next record? /no/  C
* PRINT REVERSED 15 1
Record: 43    Data base: POEMS
        This poem is from:
(etc. ... )
* PRINT LPT 13
```

COMMENTS:
The first order displays the records referred to by set 15 under the current output format for data bases. This includes some "overhead" information about the record and a break point before the verse itself is displayed. The answer "C" (typing C with the key <CONTROL> depressed) at the break point interrupts the output. The next order asks for one of set 15's records to be displayed; because of the modifier REVERSED this will be the one with the highest record number (cf. the result of the first order). The last order sends the records referred to by set 13 to the lineprinter under the current format (the break point will be ignored). If TEXTFILE is used instead of LPT then the output will be to a disk file.

```
* FORM 3
* PRINT 15 1
Record: 37    Data base: POEMS

        'Twas brillig, and the slithy toves
          Did gyre and gimble in the wabe:
        All mimsy were the borogoves,
          And the mome raths outgrabe.

* OPEN wonder mirror
```

COMMENTS:
The order * FORM 3 changes to format 3, which just displays the verse; as can be seen from the next order. The last order reopens WONDER and MIRROR.

```
* ch=num
Set 16: 375 hits
* RESTRICT 16 1 TO 2
Set 17: 53 hits in 53 records
* STATISTICS 17
Set 17:          53 hits in 53 records
Minimum:         1.0000000E+00
Maximum:         2.0000000E+00
Mean:            1.4339621E+00
Standard dev.:   5.0036264E-01
* PRINT FILE 17
Name of outfile: onetwo.ch
New file ONETWO.CH created
```

COMMENTS:
Set 16 refers to all records where "num" with search type CH occurs (i.e. to all records, as the order * STATUS ALL will tell you). Each occurrence is associated with a numeric value (the number of the chapter whence the record is taken). The RESTRICT-command creates that subset of set 16 whose associated values are 1 or 2, while the next order computes simple statistics for these values (the example is somewhat artificial, but there are few numbers in the books). The modifier FILE causes the records referred to by set 17 to be copied onto a new 3RIP-file. Give the DIR-command to see what names were assigned to its records.

```
* EXIT
Search to be restarted later? /no/ yes
File name: test
TEST is saved
Elapsed time: 1:36:49   Runtime: 0:00:19   XCS: 1074
Disk reads: 5248,  writes: 1092

.r 3RIP
(etc. ...)
* RESTART test
* GETSYM boole symbol
```

COMMENTS:
Leaving 3RIP we save the current sets in a file called TEST as we wish to continue the search later. In the next 3RIP-session (giving LEWIS as password) our first order is to restart with the saved sets from the last session. The GETSYM-command activates our saved symbol set (it is done here just to demonstrate its effect; it has nothing to do with the RESTART-command).

```
* LIST ALL
Set      Size     Order
  1        12     TEXT=PRESENT + PRESENTS
  2         3     TEXT=BIRTHDAY W1 1
  3         3     TEXT=UN_BIRTHDAY W1 1
  4         1     3 - 2
  5         1     ( TEXT=UN_BIRTHDAY - BIRTHDAY ) W1 ( PRESENT + PRESENTS )
  6         0     4 - 5
  7        15     TEXT=PRESENTED + PRESENTLY + PRESENTS + PRESENT
  8        15     TEXT=PRESENT# (naming set 7)
  9         2     TEXT=GRIN WS CAT AND PERS=CHESHIRE_CAT
 10         1     9 AND PERS=SP_CHESHIRE_CAT
 11        12     TEXT=PRESENT + PRESENTS
 12         3     TEXT=BIRTHDAY W1 11
 13         4     TEXT=WABE
 14         2     TEXT=WABE
 15         2     TEXT=TWAS W1 BRILLIG W1 "AND" W1 THE W1 SLITHY W1 TOVES
 16       375     CH=NUM
 17        53     RESTRICT 16 1.0000000E+00 TO 2.0000000E+00
```

COMMENTS:
This list shows the sets now available. Check that it tallies with what was done in the last session.

End of the example.

Managerial Aspects

System requirements:

System:	DEC-10 or DEC-20
Monitor:	TOPS-10 or TOPS-20
Source language:	BLISS 10
Memory size:	50K words + 15K words per user
Disk storage space:	
system:	2k DEC-10 blocks
data bases:	(see below; An example)

Search file systems

A search file system in 3RIP consists of a number of data bases searchable in parallel. Obviously, disparate data bases should not be brought together in the same search file system. The records of a data base are hierarchically structured in fields and subfields to enable easy categorization of the contents. When the contents of the data base are made retrievable, a number of labelled search types are established, which the users employ when searching the data base. There is not necessarily a one to one relationship between a search type and a particular part of a record. Hence any given subfield in a record can be made retrievable as more than one search type and in different manners: each word can be made retrievable or all the contents of the subfield as a phrase. The following restrictions hold (for unlimited read: whatever your installation allows):

search file systems:	unlimited
data bases in a search file system:	30
records in a data base:	4M (recommended: max. 0.5M)
characters in a record:	unlimited (recommended: majority < 2.5K)
fields in a record:	511
subfields in a field:	unlimited (max. 511 for 3RIP's editor)
search types for a data base:	30
characters in a retrieval string:	40

The recommendations for the length of the majority of records and the number of records in a data base are made because it is felt that the response times may deteriorate somewhat if these figures are transgressed. However, no application has reached these limits yet and it is quite possible that no deterioration will appear. It must be stressed that the response times within 3RIP are short and not at all sensitive to the size of the search file system; the difference between the very small demonstration system (a total of 375 records) and the largest applications (see below) is negligible in this respect.

The files in a search file system

A search file system comprises the following files:

one control file
one BASE file for each data base
one inverted file for each data base
one vocabulary file

The control file governs the whole search file system. It is the responsibility of the 3RIP-manager of the search system to create the control file. This is quite easily done with the help of 3RIP's editor. It is a small file, depending on the number of data bases, search types, etc. it may vary from 25 to a few hundred DEC-10 blocks.

The BASE file is the 3RIP-file that contains the data base itself. There are two ways of creating a BASE file: either 3RIP's external data entry format, TFORM, is used or an internal data entry format for controlled input is created and used for direct entry of the records under 3RIP. TFORM requires that an ordinary machine-readable text file is supplied with a few control characters necessary to govern the assignment of the contents of the text file to the records, fields, and subfields of the BASE file. The size of a BASE file is roughly the same as that of the same contents in an ordinary text file outside 3RIP.

Searching a data base is not done in the BASE file itself, but in the inverted file of the data base. This file contains representations of all retrieval strings in the data base with posting lists to the records in BASE where the retrieval strings occur. It is created by 3RIP's utility programs, run by the 3RIP-manager of the data base. Its size is highly dependent on the manner in which the data base is made searchable. If every word in the BASE file shall be retrievable and the option to make the contents of a given subfield retrievable as several search types is utilized, then it may be quite large indeed. However, if stop lists are employed to prohibit that very common words become retrievable and if not too much double typing is used then it will be about half the size of its BASE file.

The use of search term patterns in search orders (i.e. searching with truncated search terms, word stems, etc.) presupposes that a vocabulary file exists for the search file system. It is, however, perfectly possible to run 3RIP without this feature. One vocabulary file is shared by all the data bases in a search file system for the very good reason that if the data bases are similar enough to merit parallel searching they will to a great extent share a common vocabulary. The file contains the retrievable search terms from the data bases. If left truncation is allowed for some or all search terms it will also contain rotated forms of the search terms. It is created by 3RIP's utility programs, run by the 3RIP-manager. It is very difficult to say anything about its size. It is non-existent if pattern matching is prohibited and it could be many times the size of a BASE file if injudicious rotation is employed. For normal applications a search file system of three to four data bases may perhaps have a vocabulary file a quarter to half the size of the sum of its BASE files. Note that the more data bases there are in a search file system and the larger they are the smaller the vocabulary file will be in relation to the BASE files.

Adding new records to a data base

The utility programs used to put the first batch of records in a BASE file, invert the BASE file and put its search terms in the vocabulary file are also used when new records are to be added to the data base later. Approximately the same resources are required to add one batch of 100,000 records to a BASE file and its inverted file as to start with 10,000 records and update nine times in batches of 10,000 records. The case of the vocabulary file is somewhat different. It is much to be preferred to update this file in larger batches. However, the BASE file, the inverted file, and the vocabulary file are independent in this respect. It is quite possible to update the BASE file a number of times and postpone the updating of the inverted file and/or the vocabulary file.

An example

The table below gives some figures for file sizes and required resources for two bibliographical 3RIP data bases. The DEC-installation is the dual KL10 processor at the Stockholm University Computer Centre running under 7.00/SMP.

Data base	Number of records	chars/ rec	BASE file DEC-10 blocks	BASE file CPU mins	Inverted file DEC-10 blocks	Inverted file CPU mins	Vocabulary file DEC-10 blocks	Vocabulary file CPU mins
MARC	40,000	560	38K	22	17K	15	16K	28
MECHEN	144,000	140	33K	18	17K	15	13K	23

Some 3RIP data bases

Name	Type of records	Number of recs	chars/ rec	Owner	Site
Browny	The Brown Corpus of Present-Day American English	25K	250	English Department Stockholm University	Stockholm
COM	Computerized conference msg.	1K	300	Defence Research Inst. Stockholm	Stockholm
ICIDEMO	Washington Motor Vehicle Report	.2K	800	Information Consultants, Inc. Washington, DC	Washington
IMSL	Computer program documentation	1K	500	Stockholm University Computer Centre	Stockholm
Lobby	The Lancaster-Oslo/ Bergen Corpus of British English	25K	250	English Department Stockholm University	Stockholm
MARC	Bibliographical	40K	560	Library Linkoping University	Gothenburg
MECHEN	Bibliographical	144K	140	Info. Doc. Center Royal Inst. Tech. Stockholm	Gothenburg
PNLL	Bibliographical	2K	250	Library Polytechnic of North London	London

Appendix

paralog

presents

Alice's Adventures in Wonderland

as a computerized data base for online searching with

3RIP

3RIP is a new text data base system for DEC10 and DEC20 computers.

Alice in 3RIP provides three small full text data bases. You can search them in parallel in the same way as you would do a full text search of large text files of medical records, legal texts, news dispatches, or product data, i.e. the kind of files that 3RIP is designed to handle well at a reasonable cost.

The STATUS-command gives access to information on the available data bases (provided the data base manager has written the necessary texts). This appendix contains some of the information available for 3RIP's demonstration search file system.

A decription of WONDER

The data base WONDER contains the text from Lewis Carroll's Alice in Wonderland together with some "structural" information added by Paralog. The difficulty in putting the text in a 3RIP data base was to decide where one record should end and the next begin. We tried to find natural breaks in the context, but as the average record should not be longer than roughly one page of text some breaks between records may seem unsatisfactory.

For each record has been added: the number and title of the chapter the record is taken from; the names of the creatures appearing in the record; and the names of those who speak in the record. The names do not necessarily themselves appear in the text: there the creatures may be referred to as, for example, she or it.

3RIP's standard English command vocabulary and character set are used.

There are three TYPES available for searching WONDER:

TEXT covers that part of the record which contains the actual text from the book, except for verses (the poems in the book are not retrievable in WONDER, they have been duplicated in the data base POEMS: give the order STATUS POEMS for more information). A stoplist has been used to exclude common, trivial words, but otherwise every word has been made retrievable and given a position pointer, i. e. the search commands which measure 'distances' (W1, W2,...,W5, WS, and WP) may be used. TEXT is the default type and is necessary in orders only when changing back from another type in the same order. The order:

```
* HATTER WP HARE
```

will thus identify all records where the words HATTER and HARE occur in the same paragraph.

PERS covers the names added to the record as described above. If you are looking for the records where the Duchess appears (or is referred to) then give the order:

```
* PERS=DUCHESS
```

If you take an interest in this illustrous personage as a speaker only then your order should be:

```
* PERS=SP_DUCHESS
```

for names of speakers have been given the prefix SP_ by 3RIP. Names have not been associated with position pointers; hence the 'distance' measuring commands may not be used with PERS. A name is made retrievable as a whole so that the individual words in the name EDWIN, EARL OF MERCIA are not retrievable:

```
* PERS=EDWIN EARL OF MERCIA
```

will find him, while:

```
* PERS=EDWIN
```

will not (it might find someone else called Edwin, but not the Earl).

CH covers the number and title of the chapter whence the record is taken. A title has been made retrievable in the same manner as a name under the type PERS. Hence the order:

```
* CH=PIG "AND" PEPPER
```

will find all records from the chapter of that title (note that the word AND must be quoted as 3RIP otherwise will interpret it as the 3RIP-symbol AND) while the order:

```
* CH=PIG
```

will find nothing, for there is no chapter of that title. The number of the chapter is not directly retrievable at all. We have chosen to register the number as an integer value associated with a prefix NUM. This means that the order:

```
* CH=NUM
```

will actually identify all the records in WONDER and to find those which come from a specific chapter you must further process the resulting set with the command RESTRICT. This is a somewhat artificial manner to illustrate how 3RIP deals with numeric information (wherein Carroll's text does not abound).

Both right and left truncation is possible in searches in WONDER. The type TEXT has been fully rotated at character level, i.e. any part of a word may be given in a pattern expression. For the types PERS and CH the part of the term that is given should be a whole word. For example:

* :LO:K will find LOCK, LOOK, CLOCK, and FLOCK;

* !LOCK will find CLOCK and FLOCK;

* #HOLE will find WHOLE, MOUSE-HOLE, RABBIT-HOLE, and RAT-HOLE;

* PERS=#EARL# will find EDWIN, EARL OF MERCIA and MORCAR, EARL OF NORTHUMBRIA.

Output formats for WONDER

Here follows a description and an example of two of the three output formats available for WONDER in the demonstration search file system.

The following presentation of a typical record from the data base uses output format 2. When a 3RIP-session begins this is the active format. If you wish to change back to format 2 from another format then you should give the order: * FORM 2

If you compare this output with the output of the same record in format 1 showing 3RIP's internal structure you will see that format 2 illustrates some of 3RIP's facilities for the creation of output formats:

1. explanatory text may be added to the contents of the records;

2. the margins may be changed, indentations made, and empty lines inserted;

3. breakpoints may be given (after any subfield in the records); when a breakpoint is encountered the output is interrupted and the user is asked if the output shall continue with the same record, the next record, or not at all; breakpoints are ignored if the destination of the output is not the user's terminal.

Format 2 provides for the construction of a sortkey, hence the modifier SORTED may be used when this format is active. This sortkey will be the first 10 characters from each of the first five names of persons found under the type PERS in a record.

Here follows a typical record in format 2:

```
                        WONDER-records                  page 1
Record: 25

This record is from chapter 4
   The Rabbit Sends in a Little Bill

Persons acting or referred to in this record:
   White Rabbit
   Duchess
   Mary Ann

Speakers in this record:
   White Rabbit

Skip to next record? /no/:
          It was the White Rabbit, trotting slowly
          back again, and looking anxiously about
          as it went, as if it had lost something; and
          she heard it muttering to itself, "The
          Duchess! The Duchess! Oh my dear paws!
          Oh my fur and whiskers! She'll get me
          (etc. etc.)
```

In the demonstration search file system 3RIP's default format for non data bases, STRUCTURE, is being used as data base output format 1. This is the description (with an example) of this format:

The following display of a typical record from the data base uses the output format STRUCTURE showing the physical structure of the 3RIP-record. The marker F n,m in the margin shows that the part of the record that comes between this marker and the next is stored in subfield m of field n. This information is normally useful to the data base manager only, but may help the ordinary user to interpret the connection between the physical storing of the records of the data base and the search types used to retrieve those records (as displayed by alternative 3 in the status command for data bases).

Note that this format gives all the contents of the record:

Record: 25 Data base: WONDER

F 2,1
4

F 2,2
The Rabbit Sends in a Little Bill

F 3,1
White Rabbit

F 3,2
Duchess

F 3,3
Mary Ann

F 4,1
White Rabbit

F 5,1
It was the White Rabbit, trotting slowly
back again, and looking anxiously about
as it went, as if it had lost something; and
she heard it muttering to itself, "The
Duchess! The Duchess! Oh my dear paws!
Oh my fur and whiskers! She'll get me
(etc. etc.)

A description of MIRROR and POEMS

The data base MIRROR is exactly like WONDER, but for the fact that the texts are taken from Through the Looking-Glass.

The data base POEMS contains the verses, one in each record, from Lewis Carroll's Alice in Wonderland and Through the Looking-Glass, the texts of which are found in the data bases WONDER and MIRROR, respectively.

POEMS is very similar to WONDER and MIRROR. The main differences, except for the fact that a record contains a verse from a poem instead of a piece of text, are three:

1. no stoplist is used; every word in the verse is retrievable;

2. to each record is added the number of the verse in the poem and the number of the record in WONDER or MIRROR where the poem is stored. This information is not directly retrievable;

3. there is a special type for POEMS:
   PT for the name of the poem whence the verse in the record is taken. It is analogous with the type CH.